

Hexagon Mosaic Maps for Display of Univariate and Bivariate Geographical Data

Daniel B. Carr, Anthony R. Olsen, and Denis White

ABSTRACT. *This paper presents concepts that motivate the use of hexagon mosaic maps and hexagon-based ray-glyph maps. The phrase "hexagon mosaic map" refers to maps that use hexagons to tessellate major areas of a map, such as land masses. Hexagon mosaic maps are similar to color-contour (isarithm) maps and show broad regional patterns. The ray glyph, an oriented line segment with a dot at the base, provides a convenient symbol for representing information within a hexagon cell. Ray angle encodes the local estimate for the hexagon. A simple extension adds upper- and lower-confidence bounds as a shaded arc bounded by two rays. Another extension, the bivariate ray glyph, provides a continuous representation for showing the local correlation of two variables. The theme of integrating statistical analysis and cartographic methods appears throughout this paper. Example maps show statistical summaries of acidic deposition data for the eastern United States. These maps provide useful templates for a wide range of statistical summarization and exploration tasks. Correspondingly, the concepts in this paper address the incorporation of statistical information, visual appeal, representational accuracy, and map interpretation.*

KEYWORDS: *hexagon mosaic maps, ray-glyph maps, comparison plots, bivariate maps, brushing.*

Introduction

Maps based on hexagon tessellations are seldom used, but offer numerous opportunities for representing statistical summaries. Two fundamental maps, the hexagon mosaic map and hexagon-based ray-glyph map, provide a foundation for many task-specific variations. For example, the hexagon mosaic map shows broad regional patterns for a single variable. Variations of this map bring out distribution features of the variable represented. Important application variants include comparison of two variables through the direct display of differences and through map overlays. The ray-glyph map provides a foundation for adding more information to a map. One map variation shows local estimates and their confidence intervals. A second ray-glyph variation shows the local association of two or more variables. This variation provides an alternative to two juxtaposed maps (Monmonier 1979) and color-coded bivariate maps (Olson 1981; Eyton 1984). A third variation highlights values in specific regions based on dynamic graphic subset-selection techniques or computed criteria. This paper discusses the relative merits of these proposed maps and commonly used alternatives.

An acidic deposition study of the United States provides

examples that illustrate the map variations. The study integrates measurements from several monitoring networks and addresses the substantial variation in data quality. The variables in the study include measurements on 19 chemical ion species. The current examples use quarterly summaries for the years 1982 to 1987 and focus on sulfate and nitrate deposition. Simpson and Olsen (1990a, 1990b) describe these data in detail.

A Context for Hexagon Mosaic Maps

A tessellation is an aggregate of cells that covers space without overlapping. Only three regular polygons tessellate the plane: equilateral triangles, squares, and hexagons. This paper focuses on hexagon tessellations. The phrase "hexagon mosaic map" refers to a map that uses hexagons to tessellate major areas, such as land masses.

We do not know who created the first hexagon mosaic map. Since hexagon tessellations occur in nature (i.e., in bee honeycombs), the extension to game boards, image processing, and maps seems straightforward. We conjecture that the first use occurred in the context of games. In image processing, Pfaltz and Rosenfeld (1967) state, "Other digitized image configurations are possible, for example, using a hexagonal rather than a rectangular grid, which in fact seems preferable for some applications." The image-processing history surely goes further back. While the image-processing literature has devoted significant attention to hexagon tessellations (Serra 1982), technological convenience for raster devices and computational convenience have fostered the use of rectangular pixels. Most of today's image-processing literature addresses square pixels. Thus, students new to image processing have ample opportunity

Daniel B. Carr is an associate professor in the Center for Computational Statistics, George Mason University, Fairfax, VA 22030. Anthony R. Olsen is an ecological statistics program leader at the U.S. EPA Environmental Research Laboratory and Denis White is a research geographer with ManTech Environmental Technology Inc. at the U.S. EPA Environmental Research Laboratory, Corvallis, OR 97333.

to rediscover hexagon tessellations. This situation likely is similar in other fields. The purpose here is not to claim priority for hexagon mosaic maps, which may have been discovered independently by many different people over recent decades, but rather to further explore the use of this structure in cartography.

Hexagons have at least two advantages over squares: visual appeal and representational accuracy. Carr et al. (1987) and Carr (1991) discuss the visual appeal. The construction of maps based on either hexagon or square tessellations creates visual lines. These visual lines are artifacts of the construction process and compete with data-generated patterns. The basic claim is that humans, with their sense of gravitational balance, have a strong response to horizontal and vertical lines. Thus, the horizontal and vertical lines generated by square tessellations (in standard orientation) are particularly distracting and should be avoided.

The strong visual response to horizontal and vertical lines at different scales might be questioned. At a fine-grain level, such as using fill patterns based on parallel line screens, the visual lines should be oriented at 45 degrees from horizontal to facilitate interpretation as value. Castner and Robinson (1969) note that anyone can observe this phenomenon by examining a half-tone print in a newspaper from different orientations. Might not a strong response occur at a coarser level as well?

Consider an example that addresses symbol congestion in a scatterplot. Figure 1a shows a hexagon tessellation of a scatterplot. The size of a hexagon symbol, as shown in Figure 1b, represents the relative counts of observations falling in grid cells or "bins." The largest symbol fills the highest count cell. Figure 1b is a form of a bivariate density plot designed to handle large sample sizes (Carr et al. 1987; Scott 1992). Figure 1c is similar, but it uses a square tessellation with cells having the same area as those in the hexagon tessellation. Comparison of binned plots to the original scatterplot shows that the lattice of bin centers distracts from the pattern of the data. Comparison of the binned plots against each other suggests that the nonorthogonal visual lines of the hexagon tessellation are less distracting.

The shape (edges) of the symbols contribute to the distraction in the binned plots. Plotting round symbols in the square cells helps somewhat (i.e., the sunflower symbols in Chambers et al. [1983]), but the visual equivalent of putting round pegs in square holes wastes space and calls attention to the lattice lines. Another improvement shifts the symbol for each cell toward the center of mass of points in the cell (Carr 1991). This reduces the visual emphasis on the lattice lines. The best one can do is ameliorate the visual artifacts; the hexagon tessellation provides a good starting place.

Representational accuracy also favors using hexagon cells over square cells. For bivariate densities with the necessary derivatives, Scott (1985) has shown that hexagon-based density estimates have a somewhat smaller integrated mean square error than square-based estimates. The situation of approximating a bivariate function using a cell-based step function is similar. When the bivariate function is reasonably smooth, the range of function values over a fixed area cell generally will be smaller if the cell is as compact as

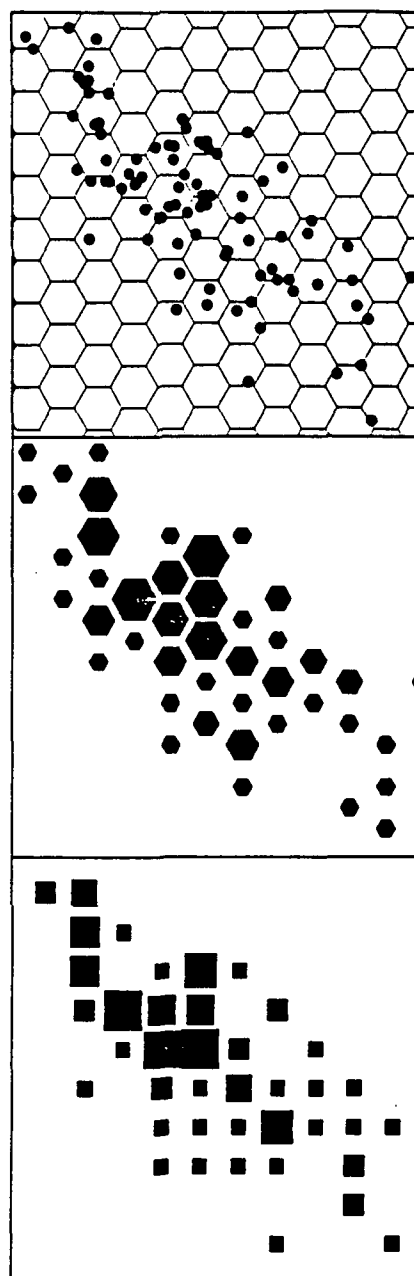


Figure 1. (a, top) Bivariate points falling in hexagon tessellation cells of a scatterplot. The bivariate points are sulfate deposition trends for 1982-87 (x-axis) and nitrate deposition trends for 1982-87 (y-axis) for sites in the eastern region of the United States. Binning consists of counting the number of points in each cell. (b, middle) Hexagon-bin bivariate density plot. The size of the hexagon symbol represents the number of points in the tessellation cell. The symbol is scaled so that the symbol representing the largest number of points exactly fills the tessellation cell. (c, bottom) Square-bin bivariate density plot. The square tessellation cells of the scatterplot are not shown. Like in Figure 1b, the size of the square symbol represents the number of points in a square tessellation cell. The horizontal and vertical lines compete for attention with the trend in the data.

possible. (One possible measure of a cell's compactness is the dimensionless second moment about the cell center, as defined by Conway and Sloane [1982]. This measure yields 0.0833, 0.0802, and 0.0796 for squares, hexagons, and circles, respectively.) Restricting the range of function values over the cell constrains the integrated mean square error. Thus, hexagon partitions generally yield better approximations than square partitions.

The function-approximation arguments in favor of hexagons are as not as strong as the visual arguments. One can construct functions for which hexagon-cell step-function approximations are not better. When the hexagon-based approximations are better, Scott's (1985) result suggests that the improvement is usually slight. Nonetheless, considerations of representational accuracy add further support to the use of hexagon mosaic maps.

One measure of success of a particular map variation is its acceptance for routine use by government agencies. A triangular grid and accompanying hexagonal tessellation has been proposed for use in the U.S. Environmental Protection Agency's (EPA) Environmental Monitoring and Assessment Program (EMAP) (Messer et al. 1991), which emphasizes probability sampling using a regular geometric arrangement of samples to achieve spatial coverage. The use of hexagonal regions is natural both for spatial sampling and data display.

EMAP's design objectives translate into a set of geometric and cartographic properties for a sampling grid (White et al. 1992):

1. Equitable spatial coverage of all environmental resources of interest
2. Random positioning to yield a probability sample
3. Equal-area sampling units to enhance precision of estimates
4. Compact arrangement of sampling units
5. Minimal correlation with any regularly spaced cultural features
6. A hierarchical structure to facilitate increasing and decreasing grid density
7. A realization of the grid on a single planar surface for the entire domain of application

The grid designed to satisfy these properties is based on a triangular array of points with a corresponding dual tessellation of regular hexagons established on the plane of the Lambert azimuthal equal-area map projection. For application in EMAP, this sampling grid has been placed on a hexagonal face of a truncated icosahedron fit to the globe (Figure 10 in White et al. [1992]). The base density grid for EMAP consists of points placed about 27 km apart and tessellation hexagons about 635 km² in area. This density represents a compromise between desired resolution and cost of sampling. The tessellation of these hexagons over the conterminous United States is shown in Figure 2.

For sampling purposes, the tessellation hexagons may be considered as strata within which point or area samples may be taken. The equal-area property assures that all points or arbitrary areas within the domain of the grid have an equal probability of being selected for sampling. For analy-

sis and display purposes, the tessellation hexagons provide a set of equal-area units that minimizes analytical and visual bias inherent in the use of arbitrary spatial units bounded by political or other features.

The Hexagon Mosaic Map and Variations

An important task for understanding sulfate deposition over the United States is to show broad regional patterns. In this case, the broad regional patterns must be constructed from irregularly sampled point data. The standard approach for this task uses a spatial estimation algorithm to obtain estimates on a regular grid. The graphical algorithms transform the regular grid estimates into maps. Our approach used kriging to produce estimates for a hexagonal grid. (Cressie [1991] discusses spatial smoothing methods, including several variants of kriging.) We then transform the estimates into a hexagon grid-cell choropleth map or hexagon mosaic map, as shown in Figure 3 (see page 271). This map distinguishes different sulfate deposition levels using color. The colors are ordered in terms of value and are light enough to allow the state boundaries generally to be visible.

The use of an equal-area regular grid on a map has interpretational advantages. In Figure 3, each hexagon covers approximately 2,670 square km. (Figure 3 has been reproduced from early studies of acidic deposition and is not an equal-area map. The area distortion in this Lambert conformal conic projection is modest, so the map has not been revised.) The total area affected by sulfate deposition, say in the range of 20 to 30 kg per hectare, can be determined simply by counting hexagons. We use this fact and determine the class intervals for the United States from the percentage of hexagons involved.

The lowest deposition class in the eastern United States covers 10% of the area and has an upper bound of 11.4 kg/ha. The next class covers 15% of the area and its upper bound is 14.8 kg/ha. Thus, 25% of the hexagons have values at or below 14.8 kg/ha. The cumulative percents used to obtain boundaries are 10, 25, 50, 75, 90, and 95. The highest deposition class covers the last 5% of the area and is shaded with the darkest color. For this application, we prefer defining classes by the percent area involved. If sulfate values determine directly the class intervals, showing the percent of hexagons involved still provides an immediate area-based interpretation.

Figure 3 itself is a variation on a basic mosaic map, since the map is presented in two parts. In this application, the big difference in values between the western and eastern regions strongly motivates the split. Notice that at least 90% of the values in the western region fall in the lowest category of the eastern region. Thus, splitting the map into two regions provides better presentation of the distribution of deposition occurring in the United States.

Alternative methods can show the information represented in Figure 3, with the leading alternatives being contour (isarithm) maps and perspective views. Colored contour maps provide the main alternative. Colored contour maps are similar to mosaic maps in several ways. Both normally involve an estimation step that produces values on a regular grid. While contouring algorithms typically

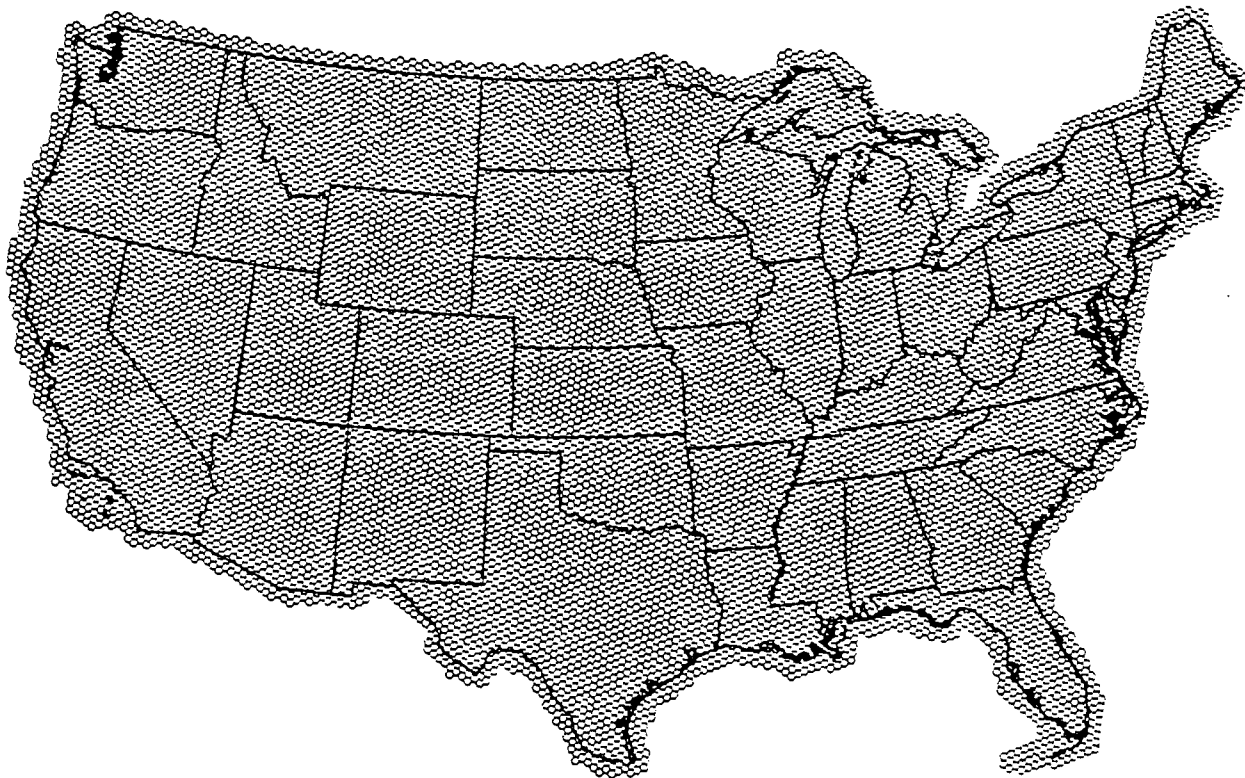


Figure 2. EMAP 635 km² nonrandomized sampling grid tessellation hexagons. The grid provides a basis for probabilistic environmental sampling.

assume a rectangular grid, it is quite possible to base both contour maps and hexagon mosaic maps on the same triangular grid of estimates. The two maps are fundamentally the same at this level of construction. Both methods represent areas and hide details concerning the amount and placement of the underlying point data. The differences between the two similar methods appear at the interpretation stage.

The hexagon mosaic map can have interpretation advantages over a contour map, because the regular tessellation suggests the use of an estimation process and facilitates thought about confidence intervals. The hexagon edges at class boundaries imply the estimation lattice that has been used. In contrast, smooth contour lines give little clue to this underlying estimation step. The value for each hexagon typically is presumed to be an estimate that represents the whole hexagon region. The value for the hexagon does not have to match the value at any particular sampling site within the hexagon. In contrast, many people interpret contour lines as precise, and knowledgeable local experts argue about the placement of a contour line relative to observed values at sampling sites.

Interpretation differences also appear in terms of confidence intervals. Analysts can easily think of the estimated value for a hexagon as having upper and lower bounds corresponding to the upper- and lower-confidence surfaces. The units for confidence bounds are data units. Pragmatically, a contour line is a sequence of line segments

whose endpoints are two-dimensional (bivariate) spatial coordinates. The notion of confidence intervals for contour lines is not well defined. Some understanding about the accuracy of a contour line can be obtained by drawing the corresponding contour lines from the upper- and lower-confidence surfaces. However, one-to-one correspondence between the points on the different lines is not guaranteed, so probabilistic considerations for contour lines do not lead to the one-dimensional simplicity of traditional confidence intervals. Consequently, the hexagon mosaic map, although less aesthetically pleasing because of its jagged boundaries, has an interpretational advantage.

Analysts often use perspective views to obtain a Gestalt impression of a surface. For rough surfaces, several perspective views may be required because part of the surface can be hidden. (See MacEachren and DiBiase [1991] for representations covering variation in continuity and abruptness.) Perspective views typically are poor for identifying the geographic location of peaks and valleys. Consequently, contour plots and perspective views often are both shown (Grotch 1983; Tufte 1991) and are best regarded as complementary, rather than competing, views.

In general, hexagons help convey the spatial structure of information. The idea of neighboring regions is clear. The notion of averaging values from neighboring regions to obtain a smoother surface estimate is straightforward and provides a reasonable introduction to more complex smoothing approaches such as kriging.

Comparison Plots

A recurrent graphical task is the comparison of two surfaces on a map (Lloyd and Steinke 1976, 1977; Steinke and Lloyd 1983; Monmonier 1990a). For example, researchers might want to compare sulfur dioxide (SO_2) emission data with sulfate (SO_4) deposition data in the attempt to better understand the deposition process. However, comparing surfaces is nontrivial. The complexity of surfaces and short-term visual memory limitations complicate the comparison of juxtaposed surface representations. Superimposed plots reduce the demand on visual memory, but restrict the form of the plot. Fishnet (perspective view) plots and translucent surfaces can be superimposed. However, this leads to a further difficulty.

Consider the simpler case of comparing two superimposed curves. Cleveland and McGill (1984) show that humans are very poor at assessing the difference between two superimposed curves. Our visual system tells us the distance between closest points on the two curves, regardless of direction, rather than computing the vertical distance between corresponding points. There is no reason to think we will do any better at comparing superimposed surface representations. Consequently, an advantageous approach computes and represents differences (and relative differences) directly. Such representations are straightforward for hexagon mosaic maps, since values are available for each hexagon.

A more limited way of making comparisons is to overlay a pair of two-class distributions, one from each surface, creating a bivariate correlation map. Olson (1981), Carstensen (1982), Lavin and Archer (1984), and Eyton (1984) discuss this type of map. Figure 4a shows the two classes from the sulfate-deposition plot. The class-interval break at 20

kg/ha was chosen because the Canadian government is concerned about values above 20 kg/ha.

The next issue to consider is which class-interval break to use from the smoothed emission data. (Smoothing facilitates area-based comparison because the emissions come from point sources, such as coal-fired plants.) Several plausible answers can be given. Figure 4b shows a class-interval break that was selected to produce the same areas as the classes in Figure 4a. Thus, attention can be focused on the location of the class differences represented in the overlay in Figure 4c. The visual impression is that the SO_2 rises and mixes with cleaner air as it moves to the northeast and is eventually deposited as SO_4 . The movement toward the northeast is only a general visual impression and does not necessarily match meteorological data. The visual impression of a much smoother deposition surface is very strong and consistent with our understanding. Use of two-class overlays provides much less information than the direct plot of differences, but this procedure can help by focusing attention on specific aspects of two surfaces.

Showing Trends and Confidence Intervals Using Ray-Glyph Maps

One purpose of the acidic deposition study was to learn about trends. For the trend study, seasonally (three-month) aggregated data for the period 1982 to 1987 are available for a number of monitoring sites. The aggregated data could be used to yield 24 seasonal maps. Comparing the sequence of 24 maps to assess trends would be a difficult visual task, complicated by the presence of seasonal variation. A direct approach shows estimated trends at each monitoring site.

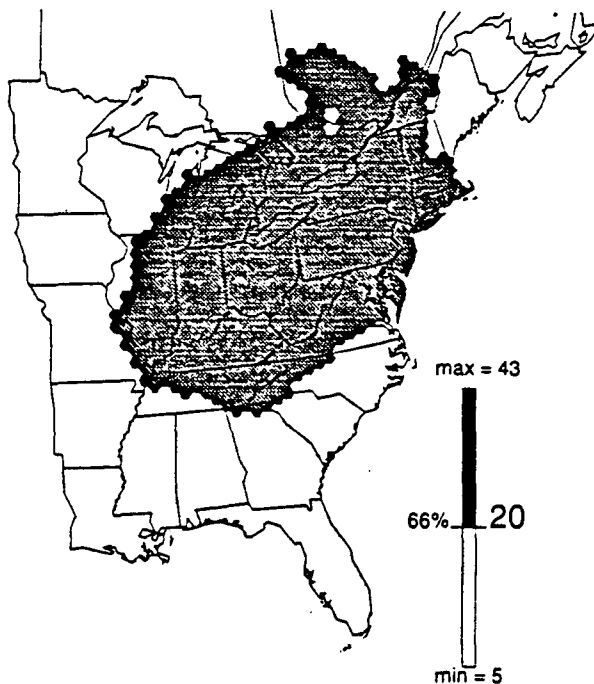


Figure 4a. Two-class map of sulfate deposition (kg/ha). The class boundary of 20 kg/ha was chosen based on policy considerations.

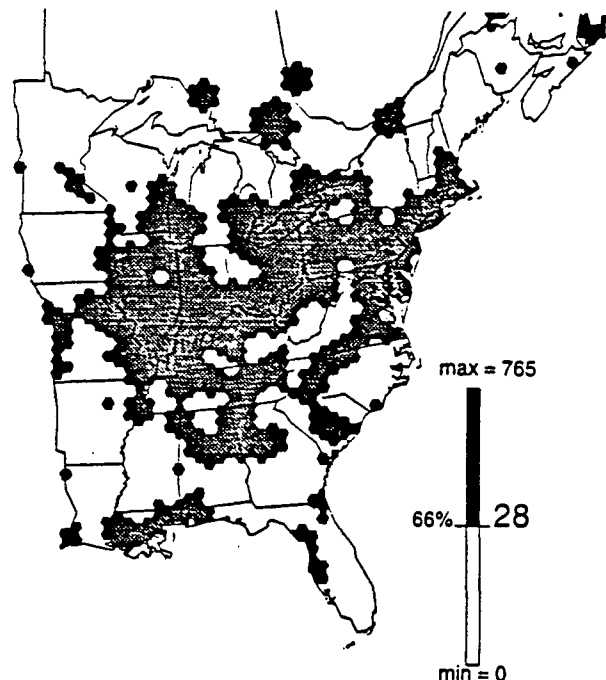


Figure 4b. Two-class map of sulfur dioxide emissions (kg/ha). The emissions surface has been smoothed somewhat, but the class boundary line is still rough. The specific class boundary defines regions with the same areas as the classes in Figure 4a.

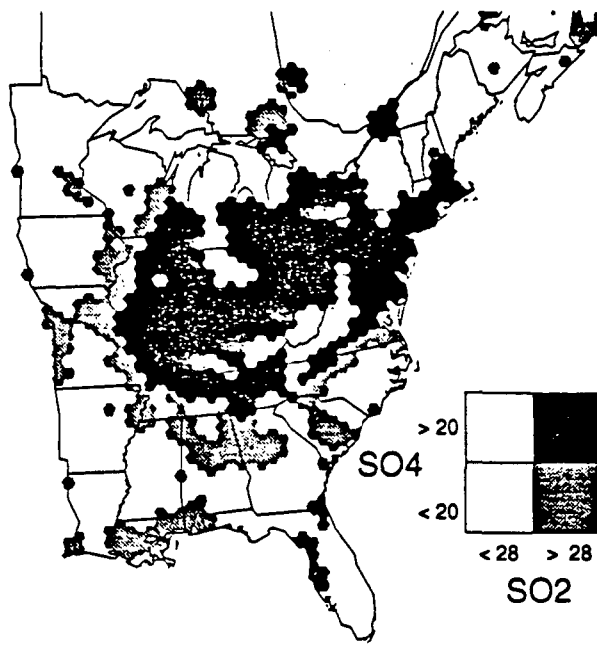


Figure 4c. Overlay of two two-class maps. Two general impressions are that the SO_4 boundary is much smoother than the SO_2 boundary, and that the SO_4 area lies more toward the northeast.

Monitoring studies often yield data of varying quality, so we use nonparametric trend methods to mitigate problems presented by poor data. In particular, we used Sen's nonparametric slope estimate and associated confidence intervals (Gilbert 1987). For each site, slope estimation proceeds by fixing a season and computing all possible pairs of slopes between years, then listing the slopes from all four seasons, and finally selecting the median from this list as Sen's median slope estimate. Confidence intervals for the estimate are based on order statistics; therefore, they are typically asymmetrical about the estimate. Computing the nonparametric slope estimates and confidence intervals for individual sites is straightforward.

The next task is to show the estimates and confidence intervals. If we want to portray only the estimates and their broad regional patterns, we could use a hexagon mosaic map. However, incorporating local confidence intervals on the map challenges us to use a different technique. The large number of sites poses a problem of graphic congestion. A solution is to aggregate the sites into hexagon regions and show only a summary slope and confidence interval for each region with data. While this introduces statistical summarization issues that deserve attention, Figure 5 illustrates the representation concept, and the octagons in the plot locate the centers of hexagon regions (not drawn) that contain site data.

Figure 5 is a ray-glyph map. The ray is composed of a line segment and a region-centered dot or polygon at the base. Ray angle encodes the slope estimate. In this case, the scaling was chosen so that a horizontal ray represents a zero-degree slope (no change). A ray straight up represents an increase of 6.6 kg/ha per year; a ray straight down represents a decrease of 6.6 kg/ha per year. In Figure

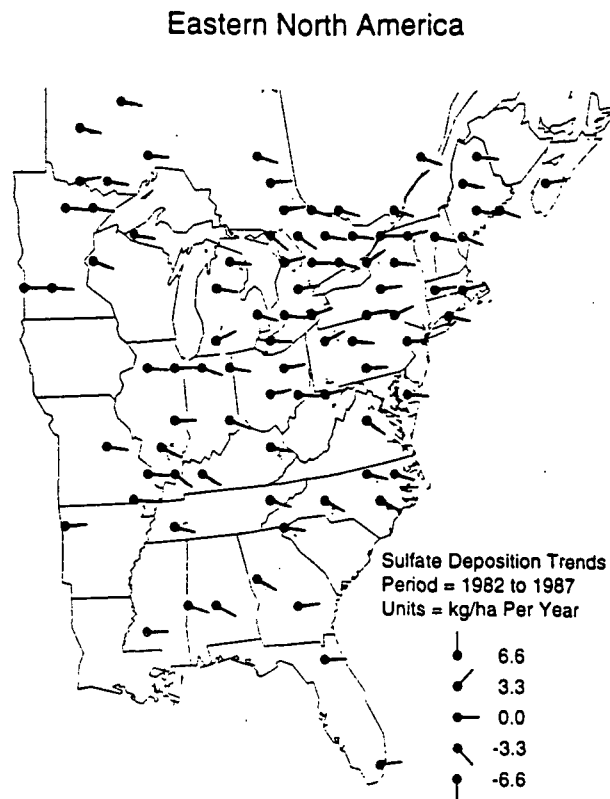


Figure 5. Ray-glyph and arc map showing sulfate deposition trends and 90% confidence intervals. The rays and arcs represent summaries for local hexagon-shaped regions. Rays with a negative slope represent a decrease in deposition. The shaded arcs covering zero suggest weak evidence concerning nonzero trends.

5, the scale upper limit for the ray is determined by the 90% confidence intervals shown as filled gray arcs; the scale lower limit is determined by symmetry. This diminishes the resolution for the estimate. (One observation with a large confidence interval was omitted for this reason.) On the plot, the majority of rays point slightly down, suggesting a small decrease in deposition over the six years. However, the majority of confidence intervals include zero, and a few might be expected to exclude zero at random since this is a multiple comparison situation. Thus, evidence for change is weak. Note also that regions without data and regions with highly variable estimates are quite evident. The ray-glyph map is effective for representing local area summaries.

Of other representations that might be considered, framed-rectangle symbols, as shown in Figure 6, provide a particularly informative alternative. Cleveland and McGill (1984) developed the framed-rectangle symbol based on their studies of perceptual accuracy of extracting the encoded information. Their studies showed that judging positions against common nonaligned scales was superior to judging circle areas, angles, colors, and other representations. Consequently, they added a frame with center ticks to provide a scale for judging bar height. The framed-rectangle symbols achieve the goal of perceptual accuracy, but other design issues also are relevant. These include addressing the

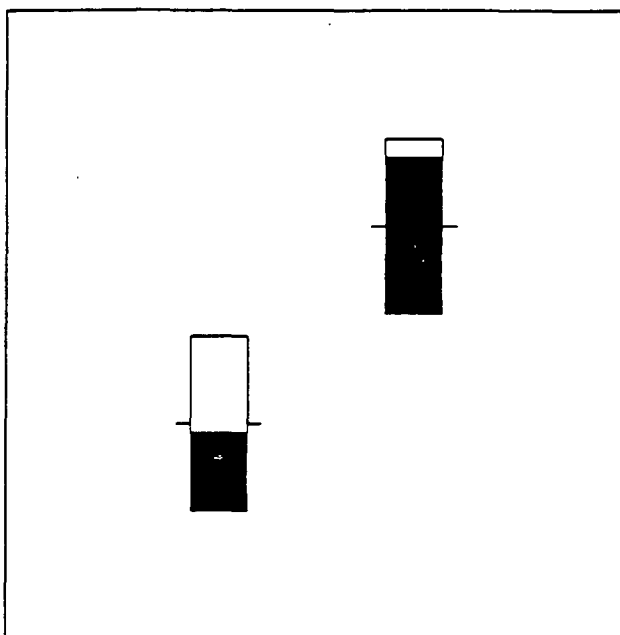


Figure 6. Two framed-rectangle plotting symbols. The frame and ticks increase the perceptual accuracy for judging bar heights. The left bar height can be assessed by comparison with the center ticks. The right bar height can be assessed by the white between the bar and the top of the frame.

recurrent problem of symbol congestion and deciding on the relative emphasis between local symbols and the other information on the map. The horizontal and vertical lines of the symbol frame draw visual attention and may interrupt the visual flow from symbol to symbol. Perceptual accuracy of extraction provides just one design criterion.

The ray glyph, with an open octagon as the base, is a line symbol and is better suited for overplotting than an area symbol like a bar. The ray glyph can be made to blend with or stand out from the rest of the map through control of line thickness, octagon size, and ray length. The octagon, with lines connecting opposite vertices, provides a visual anchor that improves the perceptual accuracy of extraction of the ray angle. When the glyph must be small, a polygon with few sides, such as a diamond, can be used as a visual anchor. In the context of hexagon tessellations, neighboring octagons also provide local scale against which angles can be judged. Positive and negative ray slopes can be assessed by looking at the corresponding octagons on the right. The ray glyph provides even greater accuracy on a hexagon mosaic map when ray scale is nested within classes that are distinguished by different colors. The ray glyph is well suited for use with hexagon tessellations and provides the flexibility to address several design objectives.

Bivariate Ray-Glyph Maps and Graphical Interaction

The representation of bivariate information using maps is a challenge. The task of interpreting side-by-side univariate maps has not proved easy, so various researchers have

proposed different methods. Monmonier (1979) discussed the cartographic cross-classification table, which is an intermediate step between juxtaposed univariate maps and bivariate maps. Carstensen (1982) and Lavin and Archer (1984) experimented with and developed continuous bivariate crosshatching maps. Wainer and Francolini (1980) and Olson (1981) have investigated the utility of bivariate color maps. Eyton (1984) has proposed additional techniques. Investigations suggest that bivariate color maps can be helpful when they have few categories and the colors are carefully selected. Bivariate ray-glyph maps (Carr 1991) provide an alternative that represents the variables with much greater resolution.

Figure 7a shows a bivariate ray-glyph map. Rays pointing to the right represent sulfate deposition trends and rays pointing to the left represent nitrate deposition trends. Dropping the confidence intervals and omitting scale considerations of zero slope and symmetry increase the resolution of sulfate deposition rays, in comparison with corresponding rays in Figure 5. The two rays of the bivariate ray glyphs generally point down or up together, so the attributes are positively correlated. With a few exceptions, the spatial change in the two trends is reasonably smooth.

Eastern North America

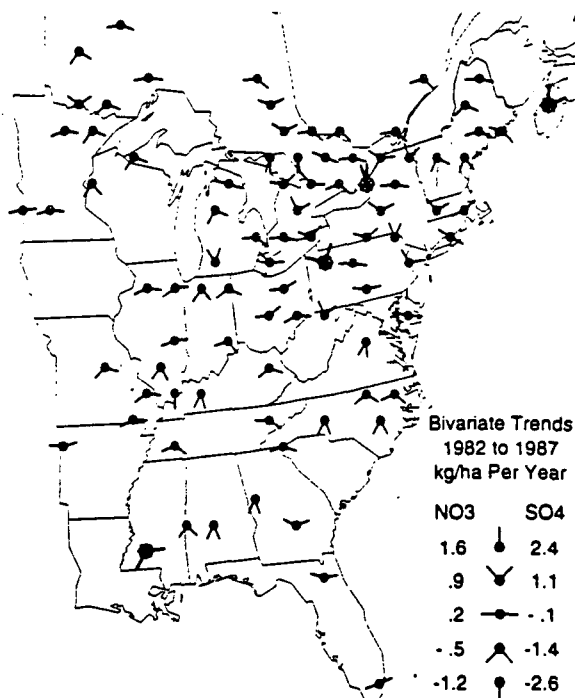


Figure 7a. Highlighted, bivariate ray-glyph map of sulfate and nitrate deposition trends. Sulfate rays point to the right, and nitrate rays point to the left. Rays pointing straight up correspond to maxima, and rays pointing straight down correspond to minima. The scaling is less than optimal for assessing zero trends, but makes assessing correlations easier. The highlighted points are determined from Figure 7b.

The bivariate ray map seems to be an effective technique for showing bivariate associations, provided the individual areas to be represented are not too small.

Having several attribute variables available for mapping introduces new possibilities. In particular, subsets can be selected from one view of the data and highlighted in another. This has been done in Figures 7a and 7b. Figure 7b shows a scatterplot of bivariate slopes. To select a set of relatively unusual points for highlighting, we computed the bivariate density for each point and chose the four lowest bivariate density points. This selection approach can easily be automated. The points have also been highlighted on the map (Figure 7a), using larger symbols with thicker lines. Alternatively, points can be selected through direct graphical interaction.

The notion of dynamic simultaneous highlighting using a mouse is often called "brushing" in the literature. McDonald (1982), Becker and Cleveland (1987), and Stuetzle (1987) provided early papers on this technique, typically in the context of attribute variable plots. Carr et al. (1987), Monmonier (1989, 1990b), Dunn (1989), and Haslett et al. (1991) have all addressed dynamic subset selection and highlighting in the geographic context. Brushing is a powerful discovery technique.

Highlighting is important in ray-glyph plots, because unusual rays may not be immediately obvious. The ideal highlight allows unusual rays to be found almost instantly. Julesz and Bergen (1983) discuss the visual phenomenon of immediate symbol location and make the distinction between rapid preattentive vision and slow attentive search. They propose the theory of textons to characterize circum-

stances under which preattentive vision locates symbols. In their theory, certain symbol characteristics are textons, and preattentive vision locates regions of rapid change in texton density. Textons include angle, color, line length, line thickness, number of endpoints, and so on. If all but a few rays had the same angle, the texton gradient would be large for anomalous rays, and the anomalous would be immediately located. However, the variety of angles shown in a ray-glyph plot reduces the gradient for unusual angles. Consequently, we use textons of line length, thickness, and endpoints to distinguish unusual rays.

New tools pose new problems. Pregibon (1989) warns of proliferating analysis alternatives. Analysts often focus on different patterns and select different subsets through brushing. When brushing produces a surprising insight, it has served its purpose. However, the typical situation is that many interesting patterns appear and the analyst wonders which, if any, are statistically unusual. Scott et al. (1978), Carr et al. (1987), and Stuetzle (1991) begin to address the statistical-assessment issue using likelihood ratios, local probabilities, and odds plots, respectively, but not in the context of geographic correlation. Brushing (yet another analysis alternative) provides the challenge to develop tools that help the analyst weed out unfruitful paths.

Opportunities and Challenges in the Larger Context

The display of quantitative information on maps is increasingly common in scientific studies. Four general factors drive this increase. First, modern data-collection systems that include satellite monitoring have made data available on an unprecedented scale. Second, the computing revolution has made it feasible for many research communities to store, retrieve, and display data. Third, visualization tools are now accepted as a legitimate part of the scientific reasoning process. Fourth, the priorities of the scientific community are shifting toward problems related to the impact of humanity on the finite earth. Thus, the scientific community is increasingly able and motivated to display information on maps.

As computing capabilities increase and disciplines begin to interact through overlapping computing capabilities and common concerns, opportunities arise to review established methods of integrating methodologies. This paper suggests rethinking the use of square tessellations in maps. Modern computing power and high-resolution plotting capabilities have taken away some of the traditional arguments that favor square tessellations. This shifts the balance toward visualization and accuracy arguments that favor hexagon tessellations.

This paper also provides a new attempt at integrating the statistical concepts of confidence intervals and bivariate correlation with mapping techniques. More integration of statistical concepts and mapping is needed. To learn about the planet earth, we need to apply sound experimental designs, appropriate spatial and temporal estimation models, powerful geographic information systems, improved visualization methods, and integrated software. This will not happen unless we collaborate across disciplines and com-

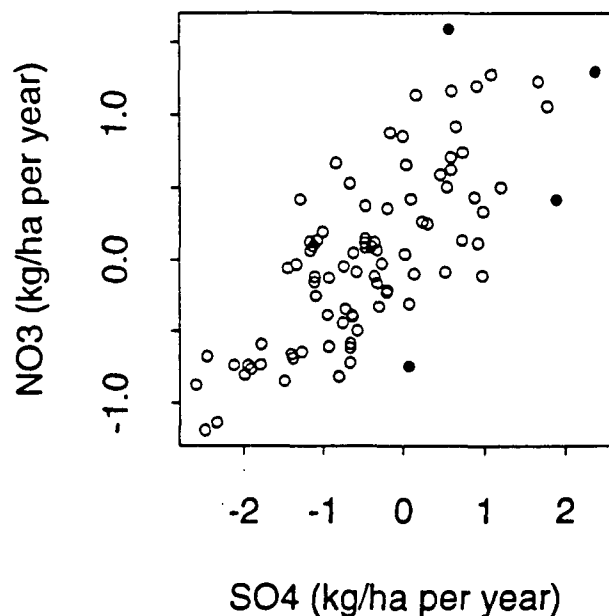


Figure 7b. Scatterplot showing correlation of sulfate and nitrate trends. The bold points have a greater distance to their respective nearest neighbors than other points, and so have been selected for highlighting. Points selected for highlighting in this way often appear less unusual in a spatial context.

municate with the public. The computing world and the rise of scientific visualization provides us with unprecedented opportunities. The pressing needs of humankind on a finite planet provide us with unprecedented challenges.

ACKNOWLEDGMENTS

The authors would like to thank Jeanne Simpson, who played an active role in the data preparation and summarization; Kevin Adams, who helped develop software for the initial hexagon mosaic maps; and the reviewers for their helpful comments. The research described in this article has been funded by the National Science Foundation under grant no. DMS-9107188 and by EPA through contracts 68-C8-0006 to ManTech Environmental Technology Inc. and 68-C0-0021 to Technical Resources Inc. This paper went through the U.S. EPA's peer and administrative review and was approved for publication.

REFERENCES

- Becker, R.A., and W.S. Cleveland. 1987. "Brushing Scatterplots." *Technometrics*, vol. 29, pp. 12-142.
- Carr, D.B. 1991. "Looking at Large Data Sets Using Binned Data Plots." *Computing Graphics in Statistics*, A. Buja and P. Tukey (eds.), pp. 7-39. New York: Springer-Verlag.
- Carr, D.B., R.J. Littlefield, W.L. Nicholson, and J.S. Littlefield. 1987. "Scatterplot Matrix Techniques for Large N." *Journal of the American Statistical Association*, vol. 82, no. 398, pp. 424-436.
- Carstensen, L.W., Jr. 1982. "A Continuous Shading Scheme for Two-Variable Mapping." *Cartographica*, vol. 19, pp. 53-70.
- Castner, H.W., and A.H. Robinson. 1969. *Dot Area Symbols in Cartography: The Influence of Pattern on Their Perception*, Technical Monograph CA-4. Bethesda, Maryland: ACSM.
- Chambers, J.M., W.S. Cleveland, B. Kleiner, and P.A. Tukey. 1983. *Graphical Methods for Data Analysis*. Pacific Grove, California: Wadsworth & Brooks/Cole.
- Cleveland, W.S., and R. McGill. 1984. "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods." *Journal of the American Statistical Association*, vol. 79, no. 387, pp. 531-554.
- Conway, J.H., and N.J.A. Sloane. 1982. "Voronoi Regions of Lattices, Second Moments of Polytopes and Quantization." *IEEE Transactions of Information Theory*, vol. 28, no. 2, pp. 211-226.
- Cressie, N.A.C. 1991. *Statistics for Spatial Data*. New York: John Wiley & Sons Inc.
- Dunn, R. 1989. "A Dynamic Approach to Two-Variable Color Mapping." *American Statistician*, vol. 43, no. 4, pp. 245-251.
- Eyton, J.R. 1984. "Complementary-Color, Two-Variable Maps." *Annals of the Association of American Geographers*, vol. 74, pp. 477-490.
- Gilbert, R.O. 1987. *Statistical Methods for Environmental Pollution Monitoring*. New York: Van Nostrand Reinhold Company.
- Grotch, S.L. 1983. "Three-Dimensional and Stereoscopic Graphics for Scientific Data Display and Analysis." *IEEE Computer Graphics and Applications*, vol. 3, no. 8, pp. 31-43.
- Haslett, J., R. Bradley, P. Craig, A. Unwin, and G. Wills. 1991. "Dynamic Graphics for Exploring Spatial Data with Application of Locating Global and Local Anomalies." *American Statistician*, vol. 45, no. 3, pp. 234-242.
- Julesz, B., and J.R. Bergen. 1983. "Textons, the Fundamental Elements in Preattentive Vision and Perception of Textures." *The Bell System Technical Journal (Human Factors and Behavioral Science)*, vol. 62, no. 6, pp. 1619-1645.
- Lavin, S., and J.C. Archer. 1984. "Computer-Produced Unclassed Bivariate Choropleth Maps." *The American Cartographer*, vol. 11, no. 1, pp. 49-57.
- Lloyd, R.E., and T.R. Steinke. 1976. "The Decision Making Process for Judging the Similarity of Choropleth Maps." *The American Cartographer*, vol. 3, no. 2, pp. 174-184.
- . 1977. "Visual and Statistical Comparison of Choropleth Maps." *Annals of the Association of American Geographers*, vol. 67, no. 3, pp. 429-436.
- MacEachren, A.M., and D. DiBiase. 1991. "Animated Maps of Aggregate Data: Conceptual and Practical Problems." *Cartography and Geographic Information Systems*, vol. 18, no. 4, pp. 221-229.
- McDonald, J.A. 1982. "Interactive Graphics for Data Analysis," Technical Report Orion 11. Stanford, California: Department of Statistics, Stanford University.
- Messer, J.J., R.A. Linthurst, and W.S. Overton. 1991. "An EPA Program for Monitoring Ecological Status and Trends." *Environmental Monitoring and Assessment*, vol. 17, pp. 67-78.
- Monmonier, M. 1979. "An Alternative Isomorphism for the Mapping of Correlation." *International Yearbook of Cartography*, vol. 16, pp. 77-89.
- . 1989. "Geographic Brushing: Enhancing Exploratory Analysis of the Scatterplot Matrix." *Geographical Analysis*, vol. 21, pp. 81-84.
- . 1990a. "Strategies for the Visualization of Geographic Time-Series Data." *Cartographica*, vol. 27, pp. 30-45.
- . 1990b. "Strategies for the Interactive Exploration of Geographic Correlation." *Proceedings of the Fourth International Symposium on Spatial Data Handling*, Zurich, pp. 512-521.
- Olson, J.M. 1981. "Spectrally Encoded Two-Variable Maps." *Annals of the Association of American Geographers*, vol. 71, pp. 259-276.
- Pfaltz, J.L., and A. Rosenfeld. 1967. "Computer Representation of Planar Regions by Their Skeletons." *Communications of the ACM*, vol. 10, no. 21, pp. 119-123.
- Pregibon, D. 1989. "Discussion of Regression Diagnostics with Dynamic Graphics." *Technometrics*, vol. 31, no. 3, pp. 297-301.
- Scott, D.W. 1985. "A Note on Choice of Bivariate Histogram Bin Shape," Technical Report 85-822-3. Houston, Texas: Department of Mathematical Sciences, Rice University.
- . 1992. *Multivariate Density Estimation: Theory, Practice and Visualization*. New York: John Wiley & Sons Inc.
- Scott, D.W., A.M. Gotto, J.S. Cole, and G.A. Gorry. 1978. "Plasma Lipids as Collateral Risk Factors in Coronary Artery Disease: A Study of 371 Males with Chest Pain." *Journal of Chronic Diseases*, vol. 31, pp. 337-345.
- Serra, J. 1982. *Image Analysis and Mathematical Morphology*. New York: Academic Press.
- Simpson, J.C., and A.R. Olsen. 1990a. "1987 Wet Deposition Temporal and Spatial Patterns in North America," Technical Report PNL-7208. Richland, Washington: Pacific Northwest Laboratory.
- . 1990b. "Uncertainty in North America Wet Deposition Isopleth Maps: Effect of Site Selection and Valid Sample Criteria," Technical Report PNL-7291. Richland, Washington: Pacific Northwest Laboratory.
- Steinke, T.R., and R.E. Lloyd. 1983. "Judging the Similarity of Choropleth Map Images." *Cartographica*, vol. 20, pp. 35-42.
- Stuetzle, W. 1987. "Plot Windows." *Journal of the American Statistical Association*, vol. 82, no. 398, pp. 446-475.
- . 1991. "Odds Plots: A Graphical Aid for Finding Associations Between Views of a Data Set." *Computing and Graphics in Statistics*, A. Buja and P. Tukey (eds.), pp. 207-217. New York: Springer-Verlag.
- Tufte, E. 1991. *Envisioning Information*. Cheshire, Connecticut: Graphics Press.
- Wainer, H., and C.M. Francolini. 1980. "An Empirical Inquiry Concerning Human Understanding of Two-Variable Color Maps." *American Statistician*, vol. 34, no. 2, pp. 81-93.
- White, D., A.J. Kimerling, and W.S. Overton. 1992. "Cartographic and Geometric Components of a Global Sampling Design for Environmental Monitoring." *Cartography and Geographic Information Systems*, vol. 19, no. 1, pp. 5-22.

Annual 1985-1987 Sulfate Deposition

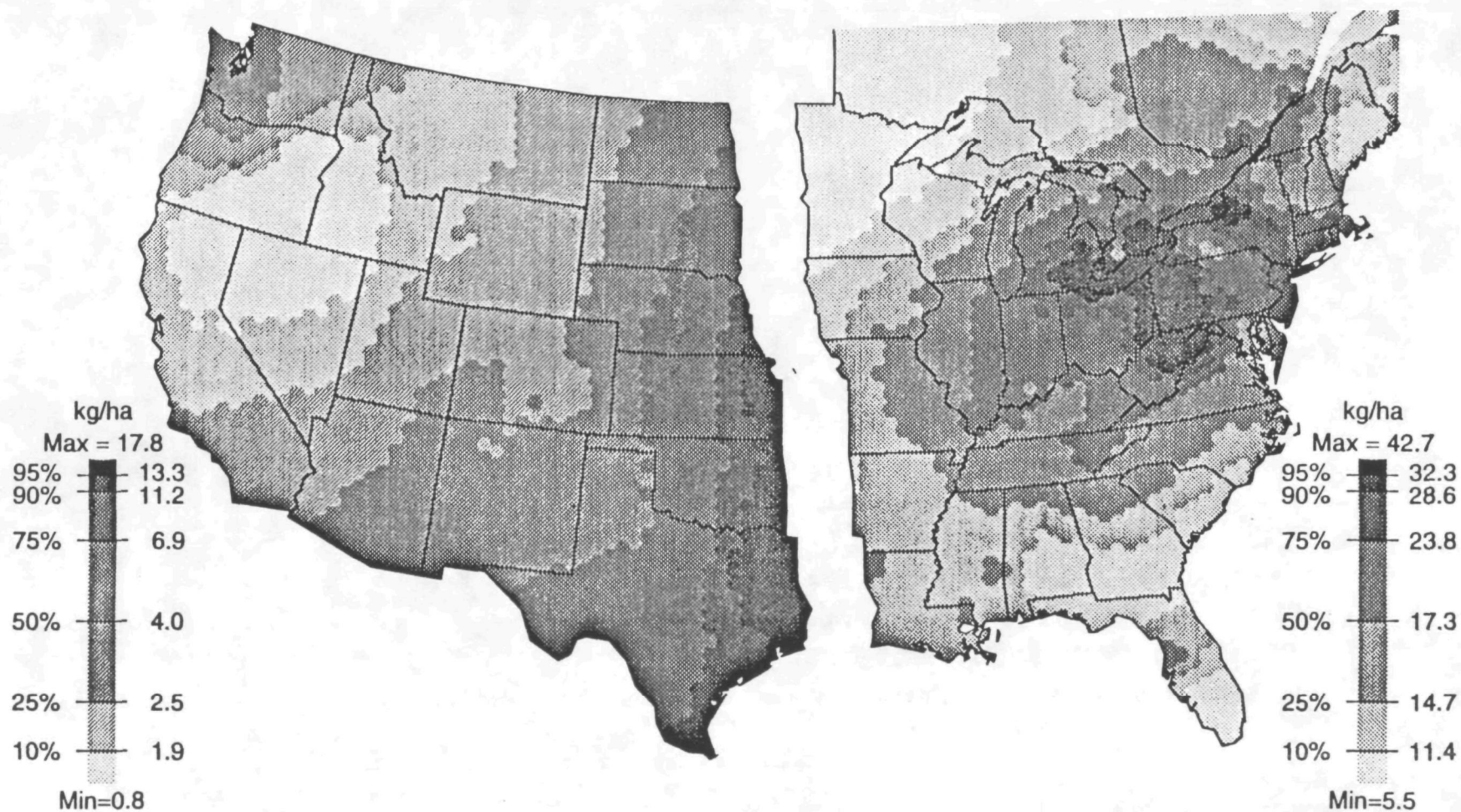


Figure 3, Carr et al. Hexagon mosaic map of sulfate deposition (kg/ha). The map is split to provide greater resolution in both the east and west. Percent of area, found by counting hexagons, determines the class interval boundaries. The hexagon edges at class boundaries indicate the hexagon cell size. The underlying estimation lattice consists of hexagon cell centers. The map is similar to a color-contour map, but suggests involvement of an estimation process.

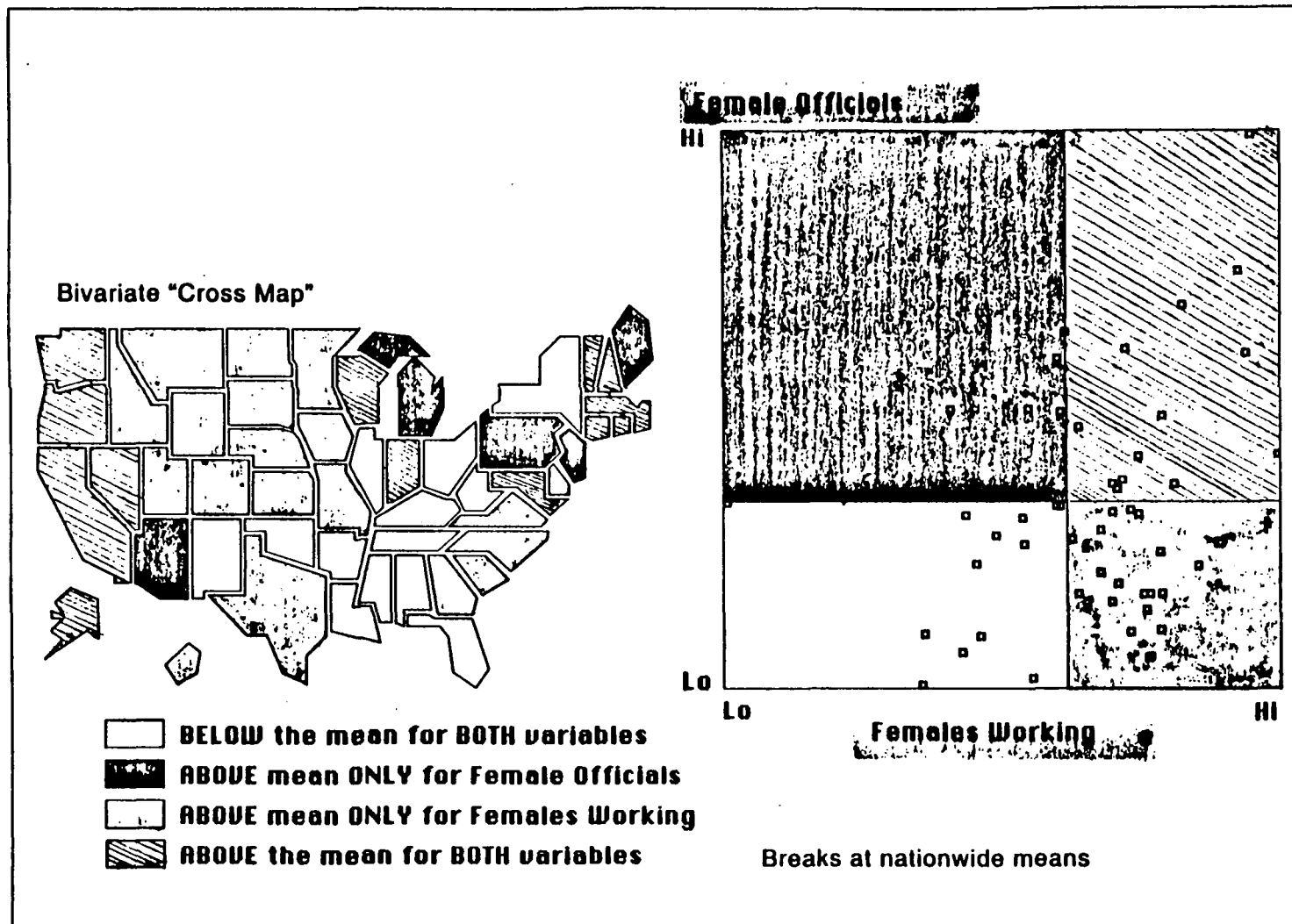



Figure 4, Monmonier. Juxtaposed cross map and scatterplot used in the correlation script's second act. Conventional key does not appear below the map until the end of the scene, after spiral variation of the category breaks (described in the text).

| TECHNICAL REPORT DATA <i>(Please read instructions on the reverse before completing)</i> | | |  |
|---|--|--|---|
| 1. REPORT NO. EPA/600/J-94/167 | 2. | 3. RI PB94-160538 | |
| 4. TITLE AND SUBTITLE Hexagon Mosaic Maps for Display of Univariate and Bivariate Geographical Data. | | 5. REPORT DATE | |
| | | 6. PERFORMING ORGANIZATION CODE | |
| 7. AUTHOR(S) D. Carr ^a , A.R.Olsen ^b , D. White ^c | | 8. PERFORMING ORGANIZATION REPORT NO. | |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS ^a George Mason Univ. ^b USEPA, ERL-Corvallis, Corvallis, OR ^c METI, Corvallis, OR | | 10. PROGRAM ELEMENT NO. | |
| | | 11. CONTRACT/GRANT NO. | |
| 12. SPONSORING AGENCY NAME AND ADDRESS US Environmental Protection Agency Environmental Research Laboratory 200 SW 35th Street Corvallis, OR 97333 | | 13. TYPE OF REPORT AND PERIOD COVERED Journal Article | |
| | | 14. SPONSORING AGENCY CODE EPA/600/O2 | |
| 15. SUPPLEMENTARY NOTES 1992. Cartography and Geographic Information Systems 19(4):228-236, 271,271 | | | |
| 16. ABSTRACT Hexagon mosaic maps and hexagon-based ray glyph maps are presented. The phrase "hexagon mosaic map" refers to maps that use hexagons to tessellate major areas of a map such as land masses. Hexagon mosaic maps are similar to color-contour (isarithm) maps and show broad regional patterns. The ray glyph, an oriented line segment with a dot at the base, provides a convenient symbol for representing information within a hexagon cell. Ray angle encodes the local estimate for the hexagon. A simple extension adds upper- and lower-confidence bounds as a shaded arc bounded by two rays. Another extension, the bivariate ray glyph, provides a continuous representation for showing the local correlation of two variables. The theme of integrating statistical analysis and cartographic methods appears throughout this paper. Example maps show statistical summaries of acidic deposition data for the Eastern United States. These maps provide useful templates for a wide range of statistical summarization and exploration tasks. Correspondingly, the concepts in this paper address the incorporation of statistical information, visual appeal, representational accuracy, and map interpretation. | | | |
| 17. KEY WORDS AND DOCUMENT ANALYSIS | | | |
| a. DESCRIPTORS | b. IDENTIFIERS/OPEN ENDED TERMS | c. COSATI Field/Group | |
| hexagon mosaic maps, ray-glyph maps, comparison plots, bivariate maps, brushing. | | | |
| 18. DISTRIBUTION STATEMENT Release to Public | 19. SECURITY CLASS (This Report) Unclassified | 21. NO. OF PAGES 8 | |
| | 20. SECURITY CLASS (This page) Unclassified | 22. PRICE | |